TURTLE GAMES

Advanced Analytics for Organisational Impact

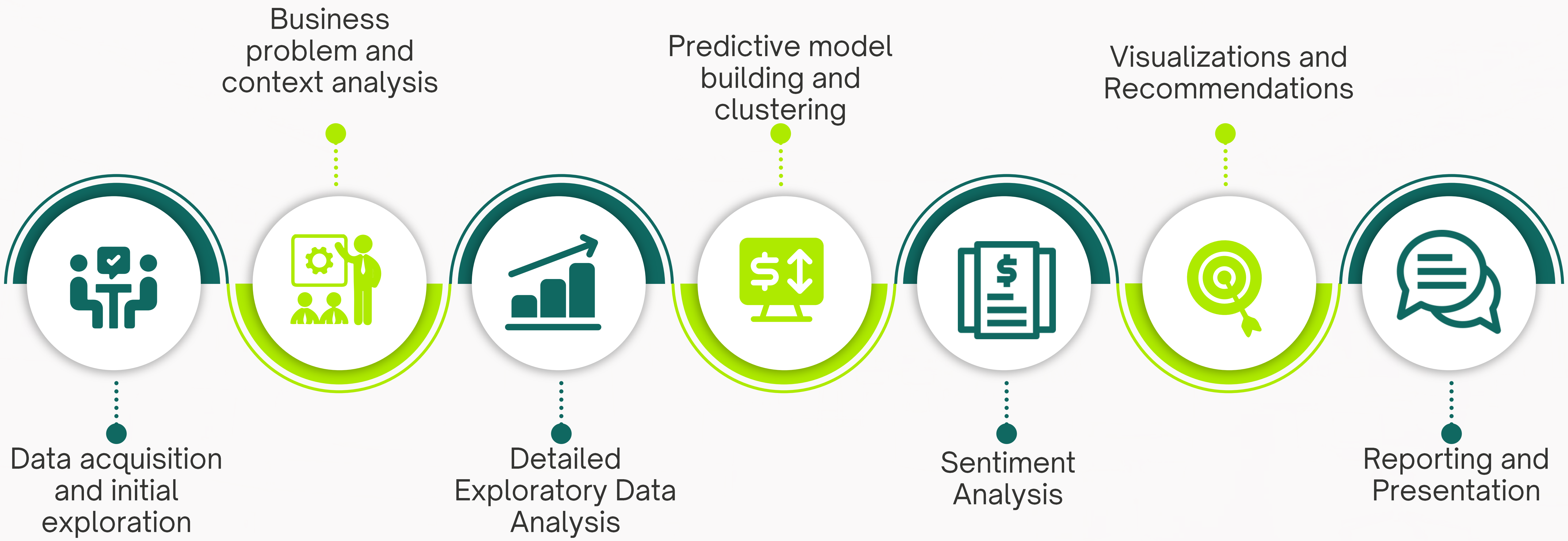
Presented by Beata Faitli
LSE Career Accelerator Student

Business Problem

Turtle Games aims to improve overall sales performance by analyzing customer loyalty points, spending behaviors, and customer sentiment from reviews. The goal of the analysis was to identify high-value customers, improve customer satisfaction, and refine marketing strategies.

Analysis Timeline

Python and R was used for a comprehensive analysis. Key stages included data cleaning, exploratory analysis, transformations and preparation for predictive analysis, customer clustering and sentiment analysis.



Exploratory Data Analysis
Patterns, Trends, and Insights

Key customer demographics

Age

Ranges between 17-72
Average age of 39 years

Income

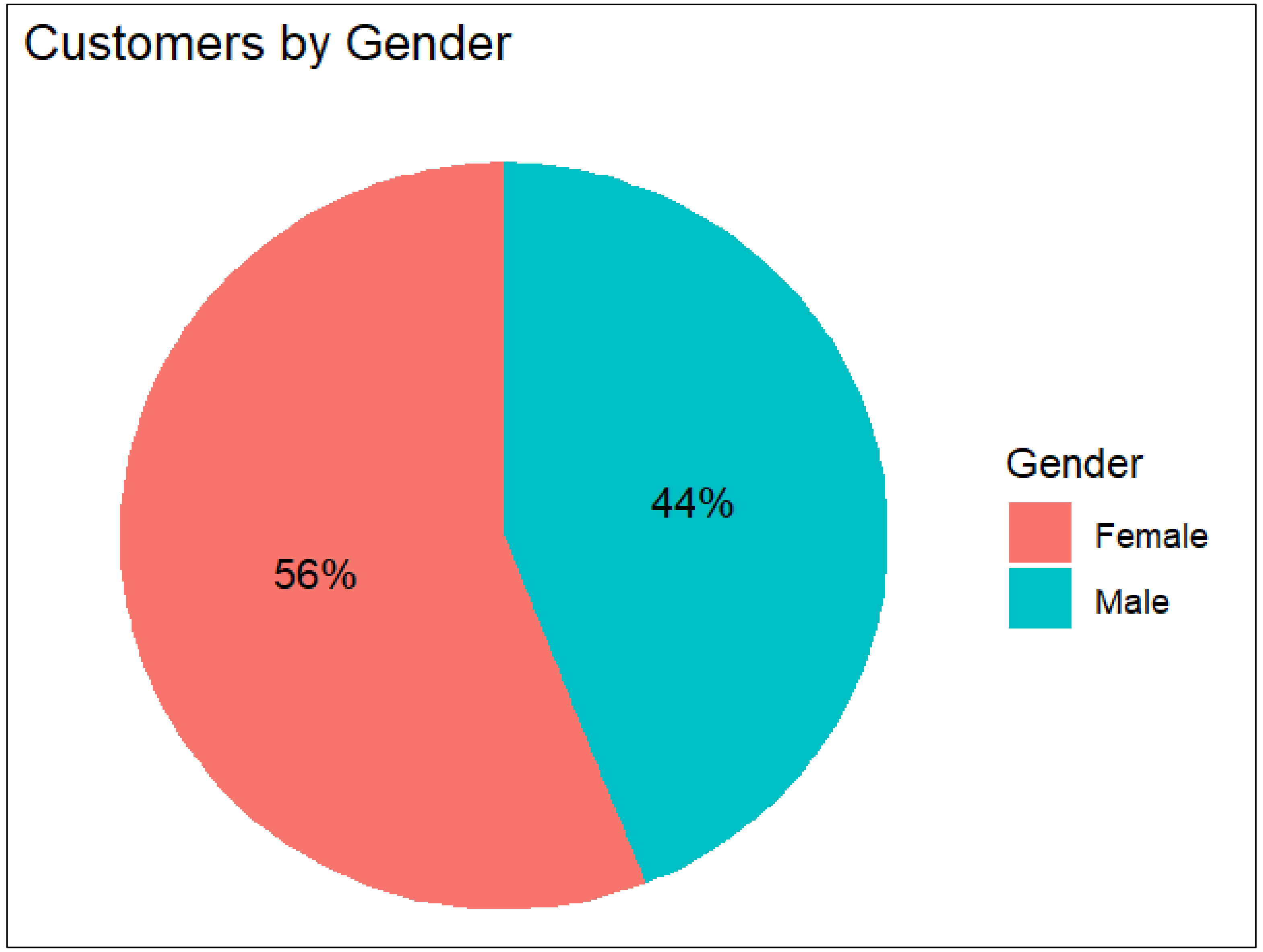
Ranges between $12.3K to $112.3K
Average income of $48K

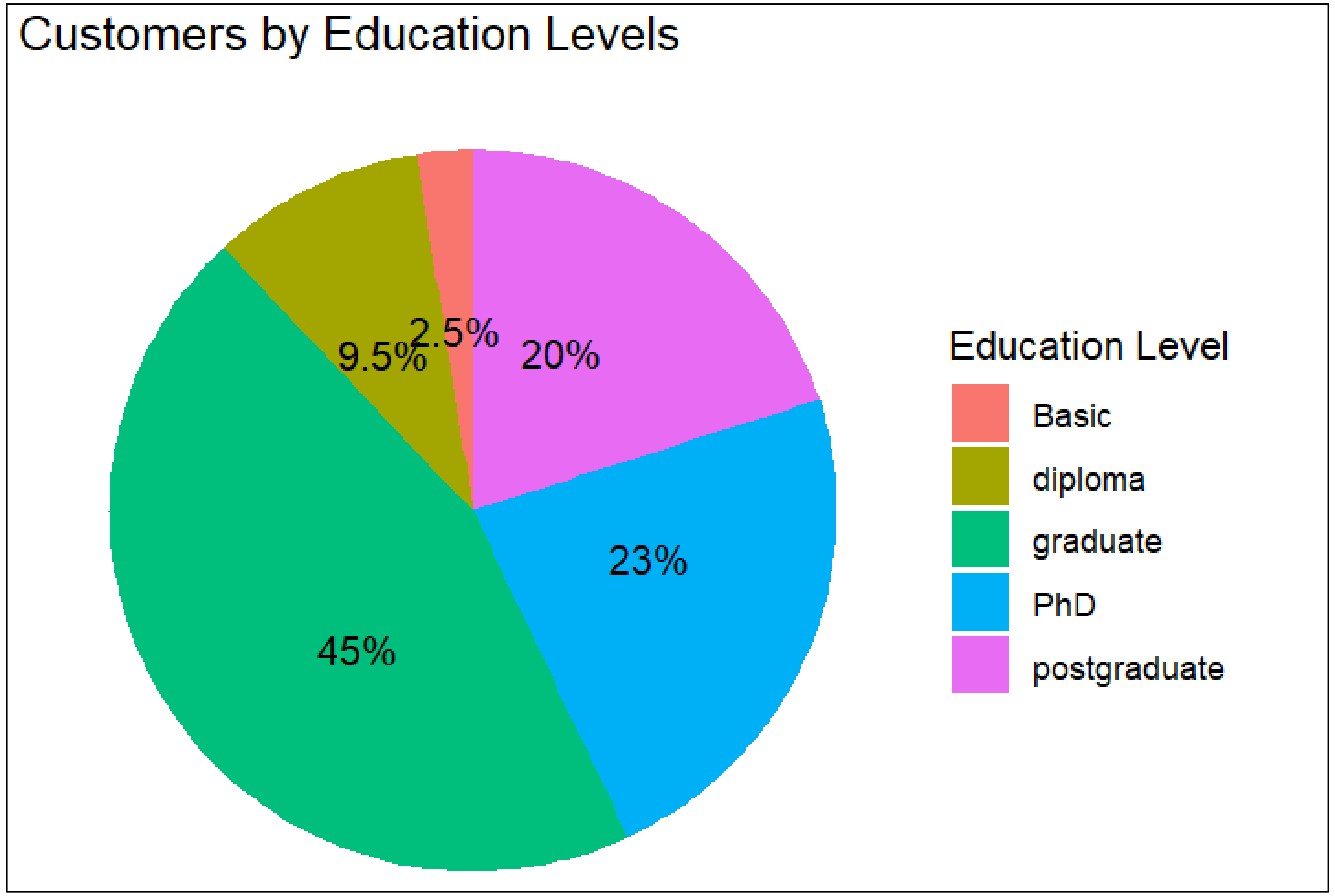
Spending Score

Ranges between 1 to 99
Average spending score of 50

Total Number of Customers: 2000

Customer Demographics Overview
Gender and Education Levels



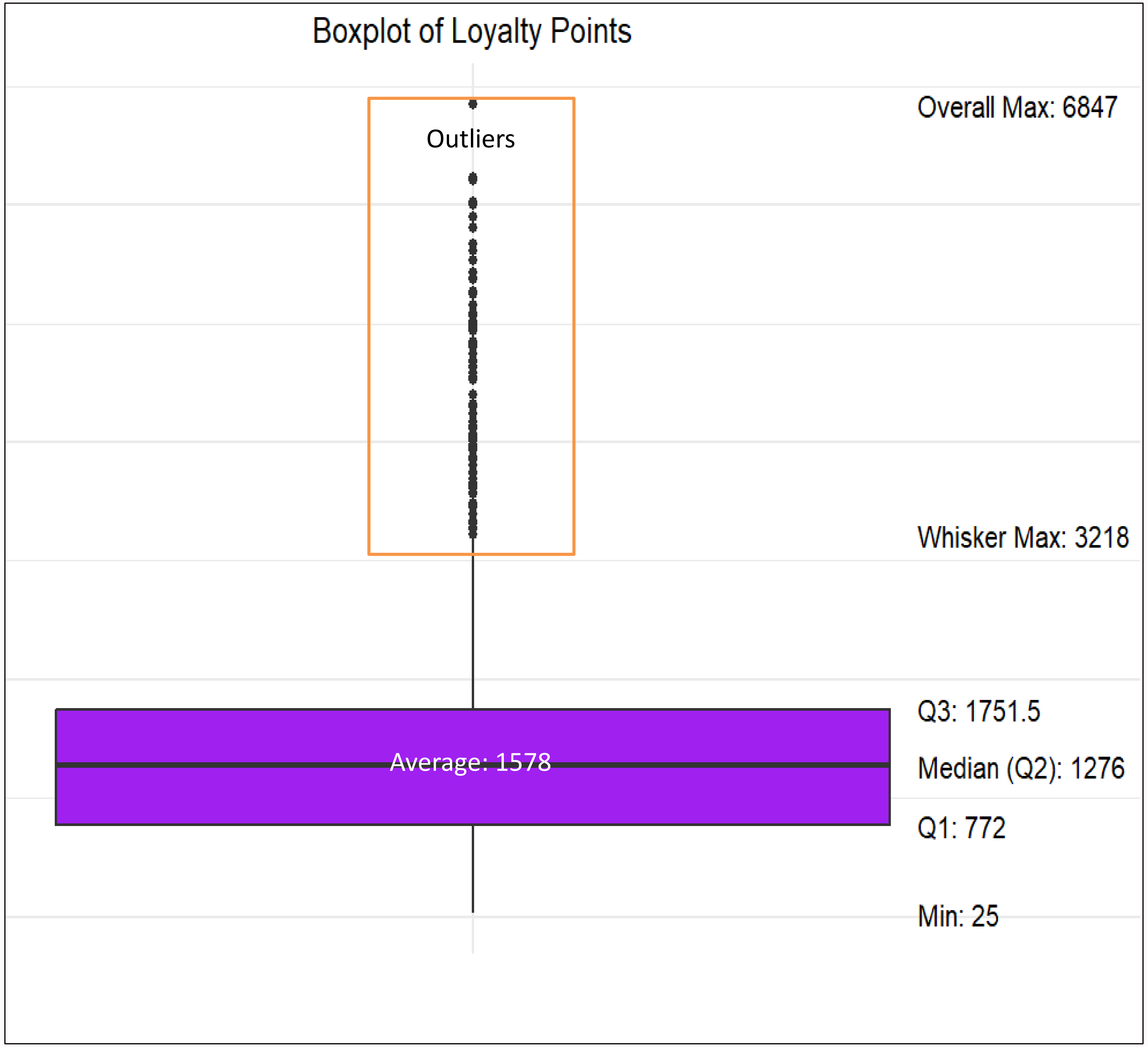


Key Insights:

- Most customers hold graduate or higher education levels, indicating an audience potentially receptive to intellectually engaging marketing.
- Higher percentage of female customers.

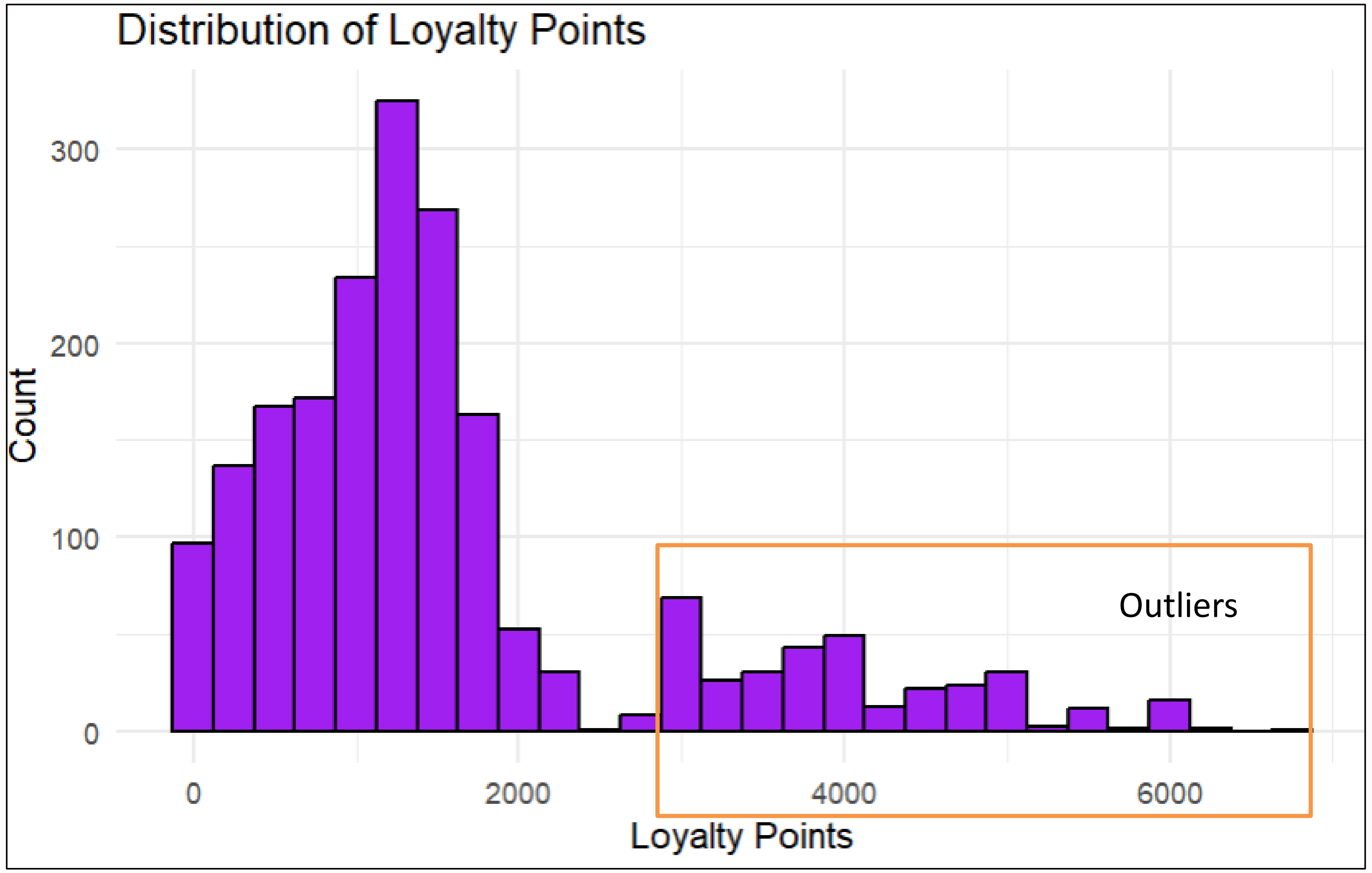
Note: Potential data quality issue - 50 customers less than 20 years old reporting PhD and Postgraduate level of education

Loyalty Points Overview: Key Outliers and Distribution



Outliers – above 3218 loyalty points

- 13% of customers
- "VIP" / high loyalty customers - Key targets for premium loyalty programs
- Excluded from certain predictive modelling to improve model accuracy



Key Drivers of Loyalty Points Accumulation

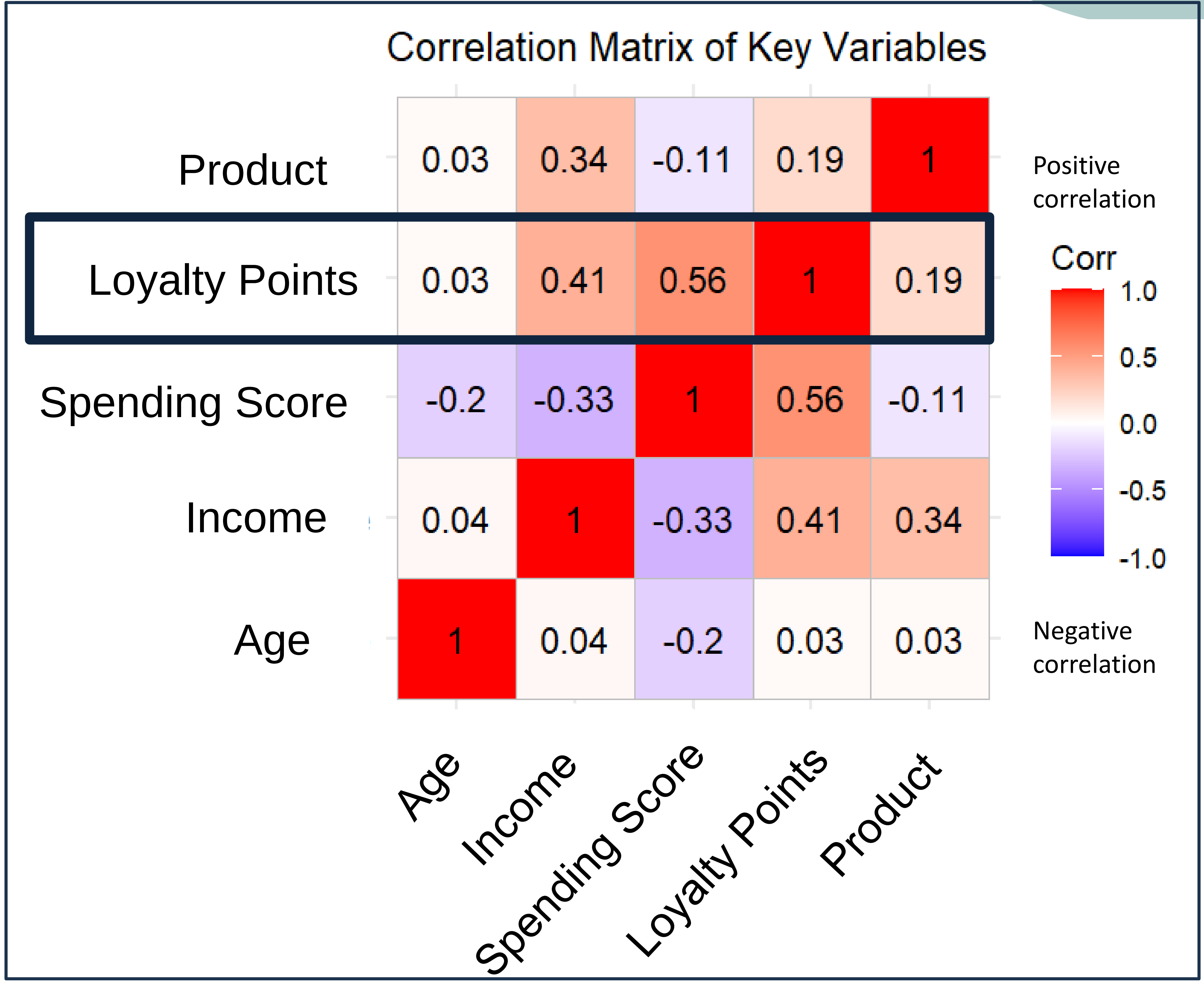
Income and Spending score

- Spending Score has strong positive (0.56) correlation with loyalty points.
- Income has moderate (0.41) correlation with loyalty points.
- Age and categorical variables, such as gender and education level, show no significant correlation with loyalty points.
- Spending score and income are also mildly positively correlated.

Recommendation:

- Incentivizing spending

Target customers with high spending scores for loyalty rewards to maximize their engagement.

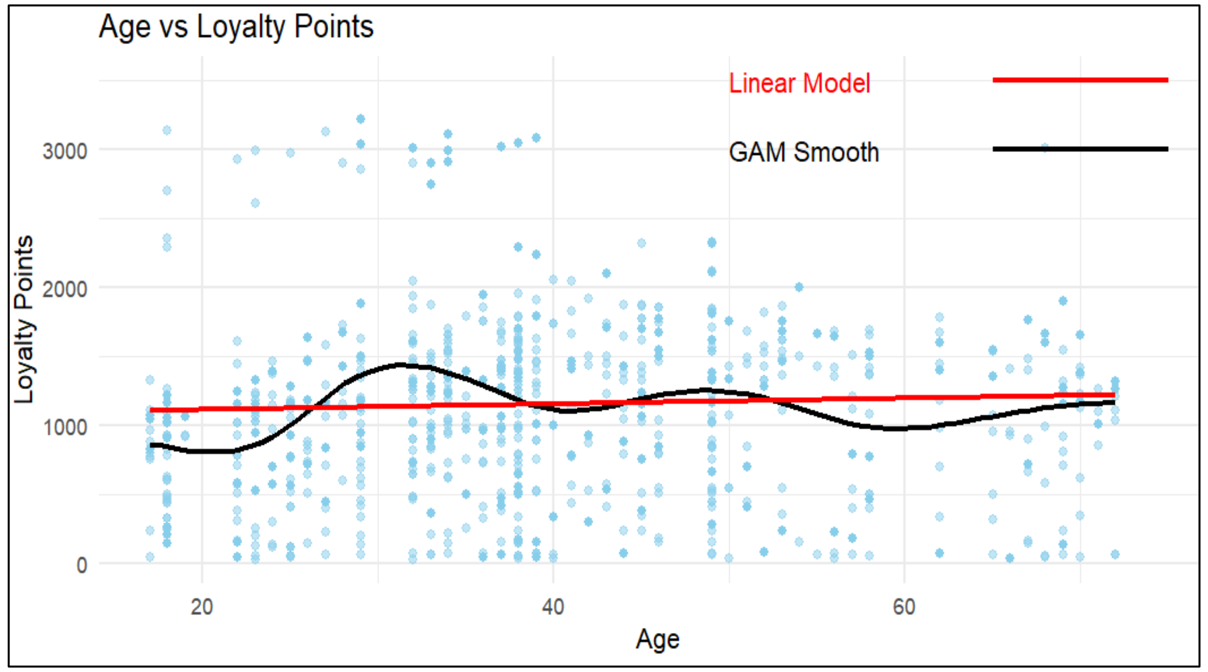


Predictive Modelling

Simple Linear Regression Modelling

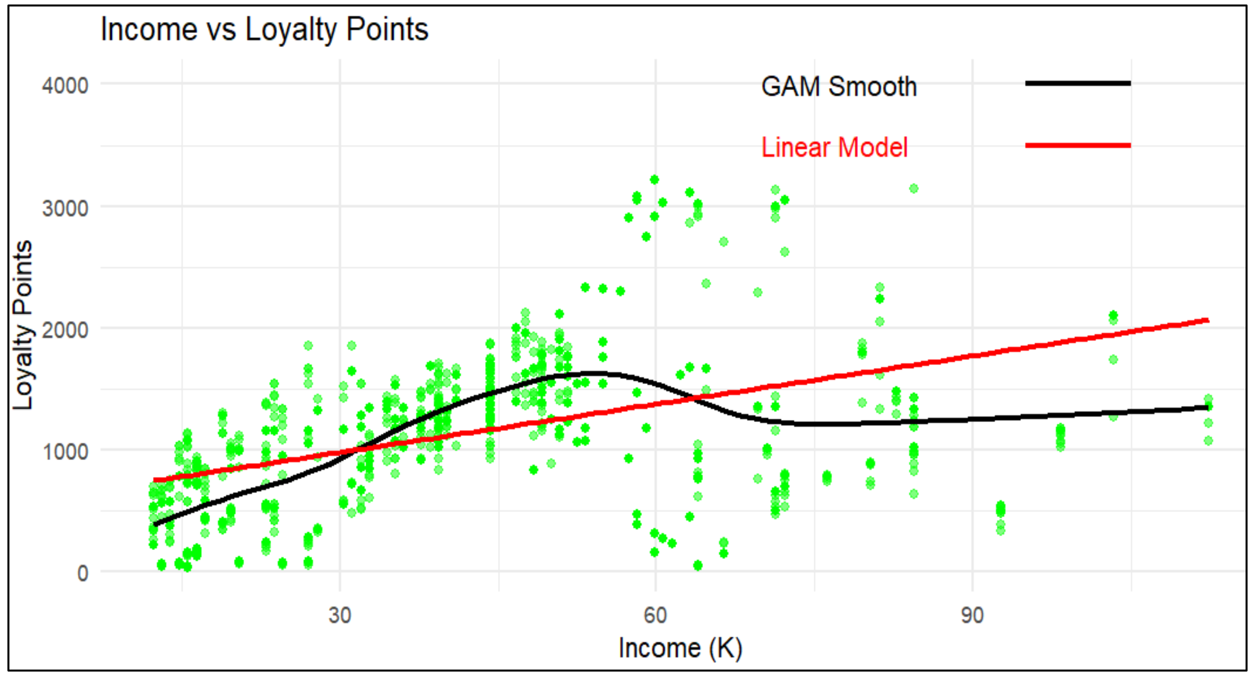
Simple linear regression models on their own were **not able to capture the complex relationships** that explains accumulation of loyalty points.
As seen previously from the correlation matrix **spending score emerges as the most significant predictor** for loyalty points accumulation.

Age



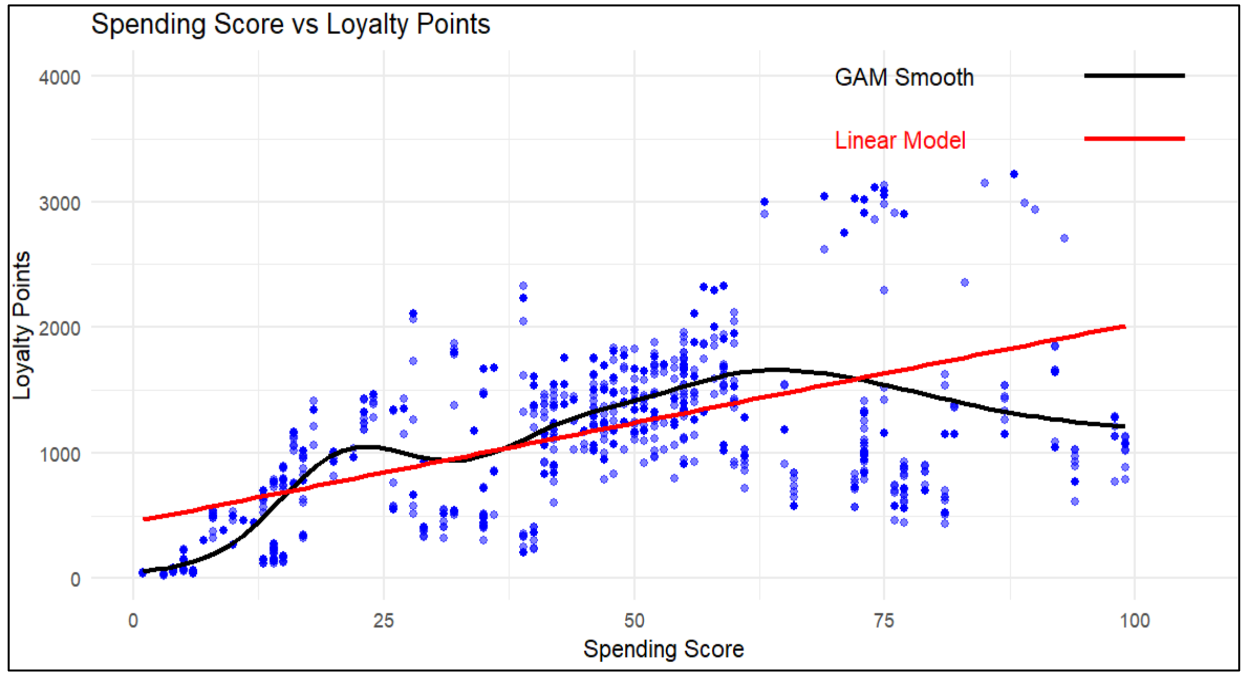
Age has no influence on Loyalty points

Income



Income explains 16% of the variation in Loyalty points

Spending Score



Spending score explains 31% of the variation in Loyalty points

Key Findings from Predictive Modelling

Refer to report for detailed model description and comparisons

Whilst single linear regression models were not suitable for the dataset. Several other models could be deployed to predict loyalty points and high-value customers, showing strong performance overall.

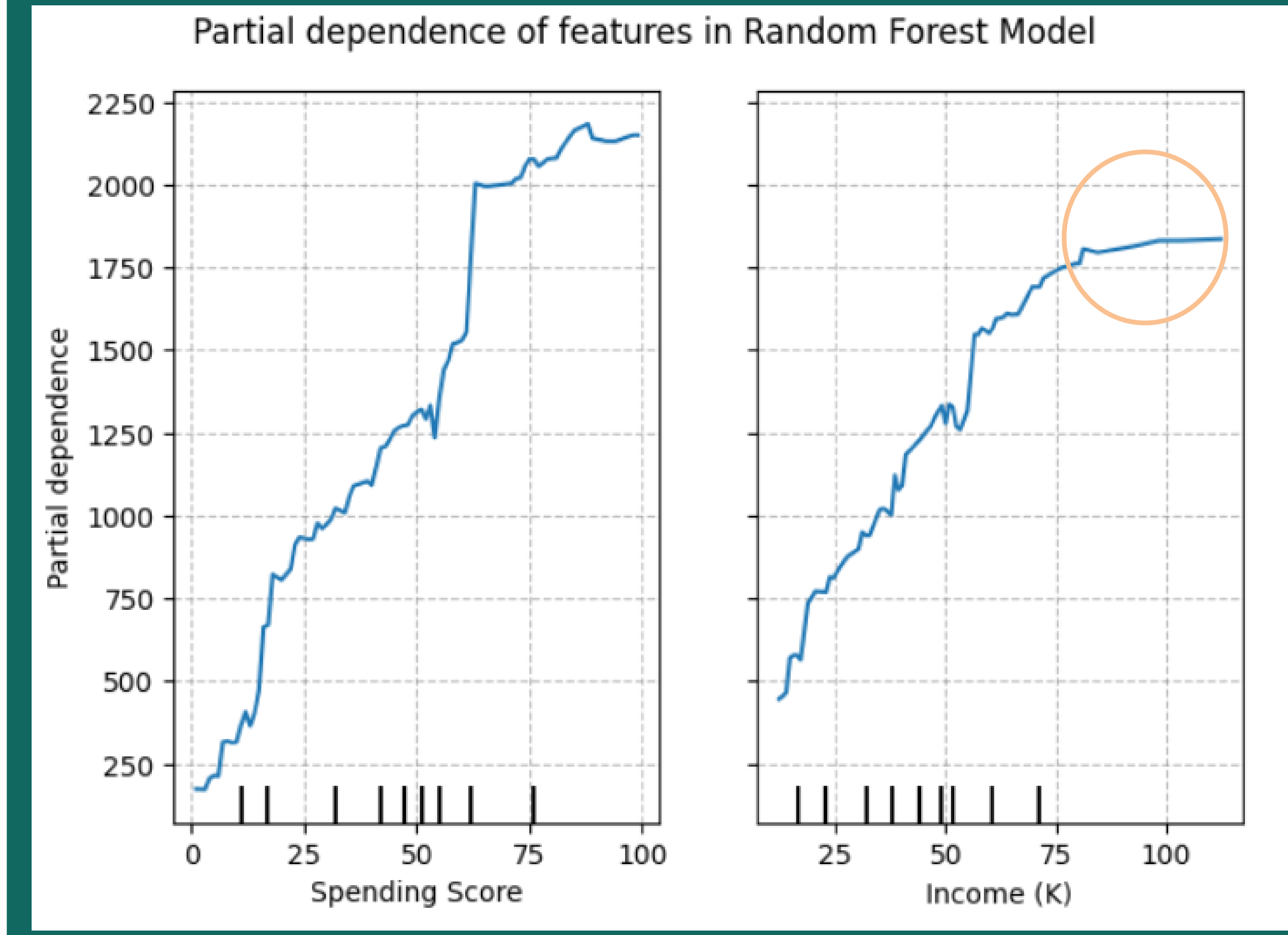
- **Key Predictors**:
  - *Income* and *Spending Score* emerged as key drivers across all models.
  - *Diminishing Returns*: Higher income levels (above $75K) showed diminishing returns on loyalty points, emphasizing the need for tailored incentives for high-income customers

- **Best Performing Models:**
  - **Multi-Linear Regression for quick insights:** Simplest model with 94% adjusted R-squared when transformations and further filtering applied.

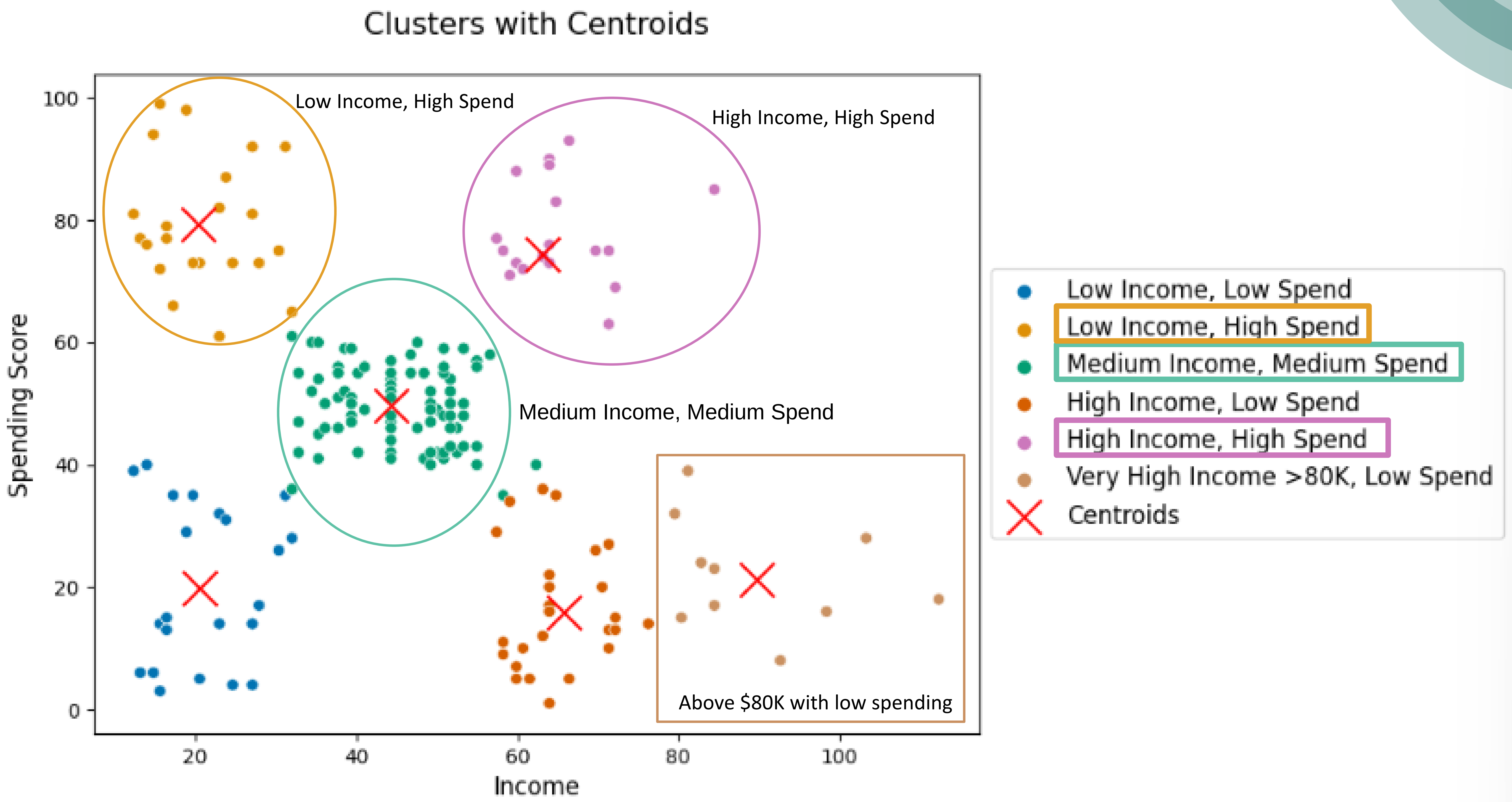
  - **Random Forest for highest accuracy**: Highest accuracy in predicting loyalty points (98%), but more complex and computationally intensive.

  - **Decision Tree for easy interpretation:** Easy to interpret, ideal for real-time segmentation with accuracy around 90% at level 5 pruning.

  - **Logistic Regression Models to predict loyal or high-value customers** using composite score: 98-99% accuracy and low misclassification rates.



Key Customer Clusters Overview



Key Customer Clusters Overview

Cluster Names	Percentage of Total Customers	Average Loyalty Points	Average Age	Top Education Level	Top Gender	Top Product	Sentiment for Top Product	Most Common Review Sentiment	Average Review Length (words)
High Income, High Spend	5	2982.0	33.0	graduate	Female	7384	Neutral	Positive	55.0
Medium Income, Medium Spend	44	1422.0	42.0	graduate	Female	399	Positive	Positive	64.0
Very High Income >80K, Low Spend	7	1278.0	40.0	PhD	Female	3711	Positive	Positive	57.0
Low Income, High Spend	16	972.0	32.0	graduate	Female	2162	Positive	Positive	50.0
High Income, Low Spend	12	711.0	41.0	graduate	Male	3158	Positive	Positive	50.0
Low Income, Low Spend	16	276.0	43.0	graduate	Female	1031	Positive	Positive	53.0

Actionable Insights:

Strategies for best (ROI) Return on Investment Groups

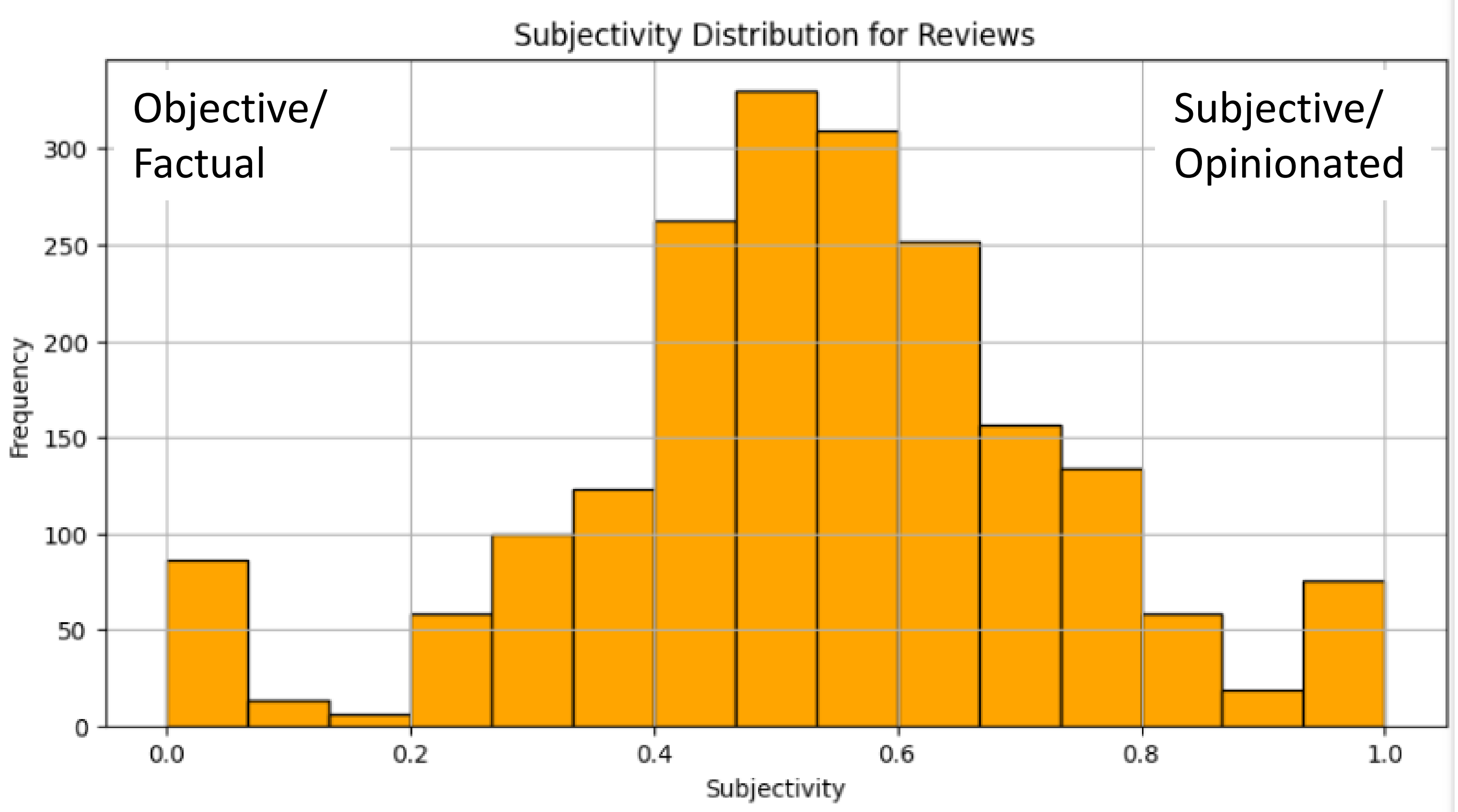
- **Improve product sentiment** for High Income, High Spend Customers
- **Increase engagement and spending** in Medium Income, Medium Spend customers given their large customer base
- **Offer personalized rewards or incentives** to the Low Income, High Spend group, as their behaviour shows high spending potential despite a lower income level.

Sentiment Analysis of product reviews

Product Review and Summary Initial insights

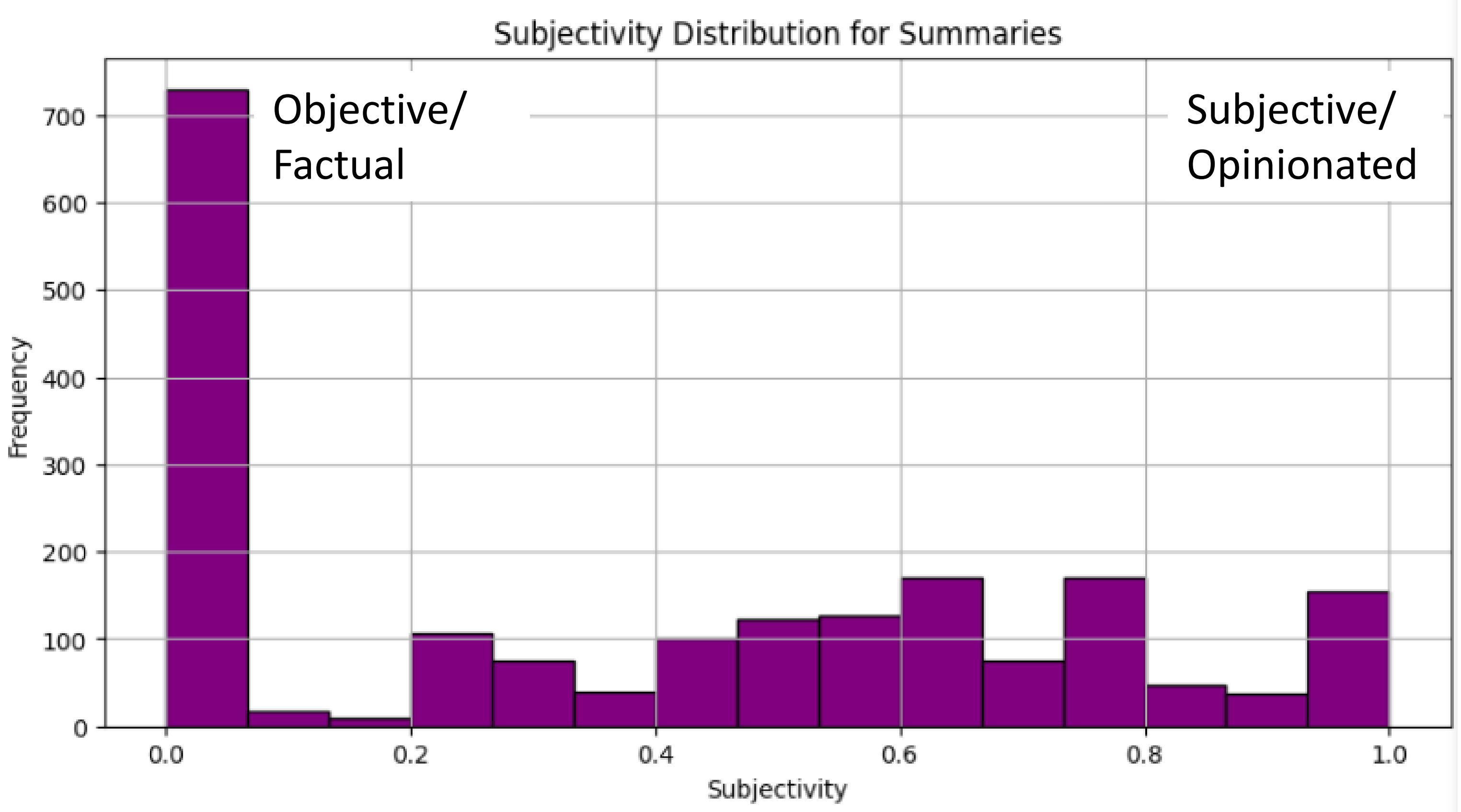
Product Reviews

- The most popular product was reviewed 11 times
- The code applied to future **reviews returns insights in minutes**
- **Average review length was 56 words.**
- Only 25% of the reviews were longer than 61 words, with the longest of 1437 words
- **Short reviews dominated every sentiment**
- Reviews showed a **broad range of subjectivity**



Review Summary

- **Approximate time to complete: 36 hours**
- **Inconsistent summarizing method:**
  - **25 %** of reviews were summaries with a **star rating** (Average review word length of the 7 words)
  - **75 %** of reviews summarized with a **short word summary** (Average review word length 71 words)
- Summaries tend to be **more objective/factual**, with a larger percentage of them scoring lower on the subjectivity scale compared to reviews.



Sentiment Analysis Using TextBlob & VADER

For more details refer to the report

VADER Analysis

VADER (Valence Aware Dictionary for Sentiment Reasoning) tool. It is suitable for short, informal text (like tweets or product reviews, or feedback with emoticons and punctuation)

Uses compound scoring for sentiment

- **Advantages:**
  - Highlights the longer positive/ negative reviews due to the compound effect of scoring
  - **VADER** scores higher in technical **accuracy** for review/summaries comparisons
- **Limitations:**
  - May over-represent sentiment in longer texts due to its compound score

TextBlob Analysis

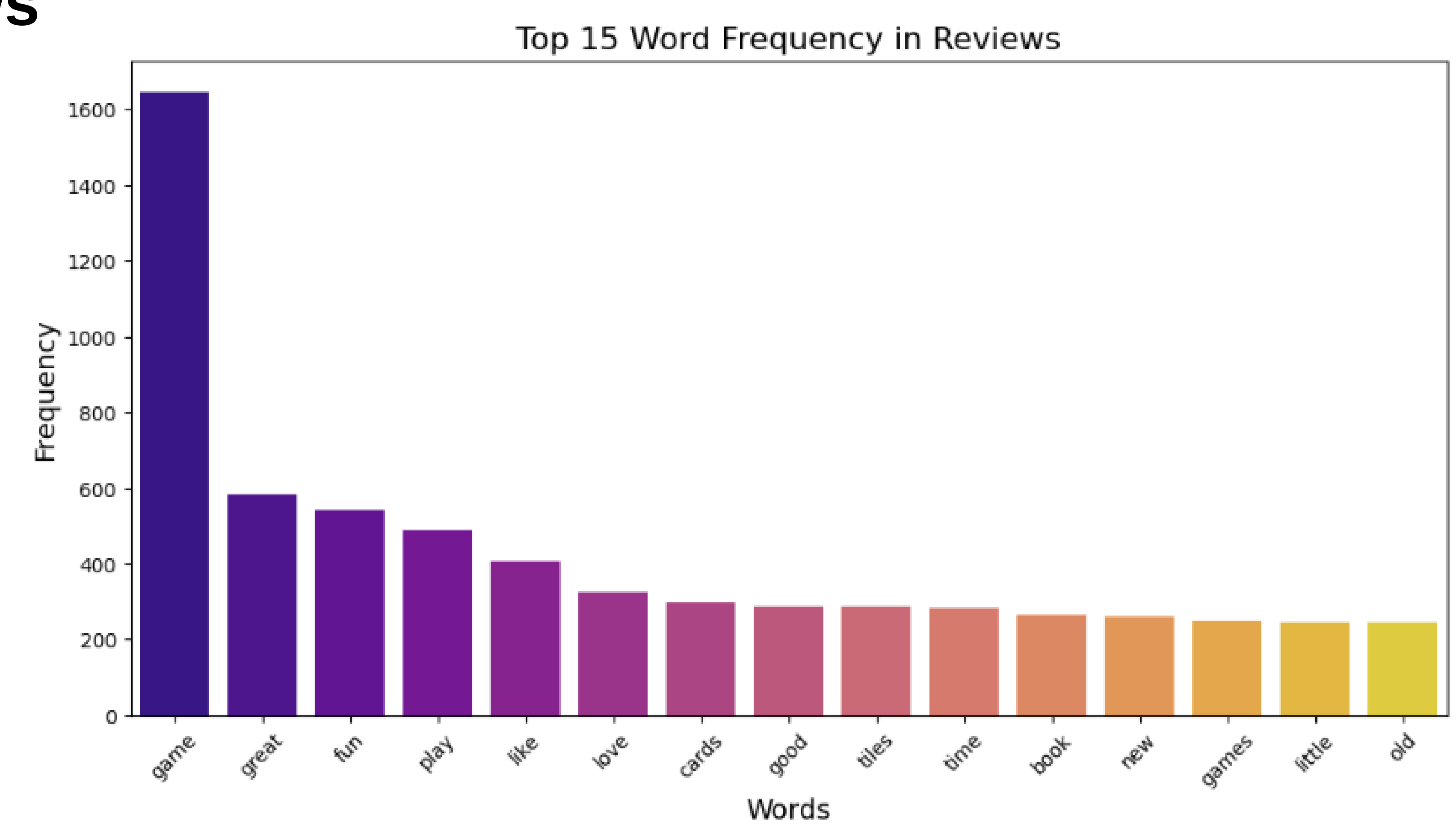
TextBlob is a simple and effective NLP tool, ideal for sentiment analysis. Effective in analyzing formal customer summaries or articles with structured sentences.

- **Advantages:**
  - Returns short quick reviews for top 20 positive/negative reviews
  - **TextBlob** gives a more **customer-aligned output** when focusing on **overall positive/negative satisfaction** as seen through CSAT.
- **Limitations:**
  - Might miss certain nuances in informal language or sarcasm.

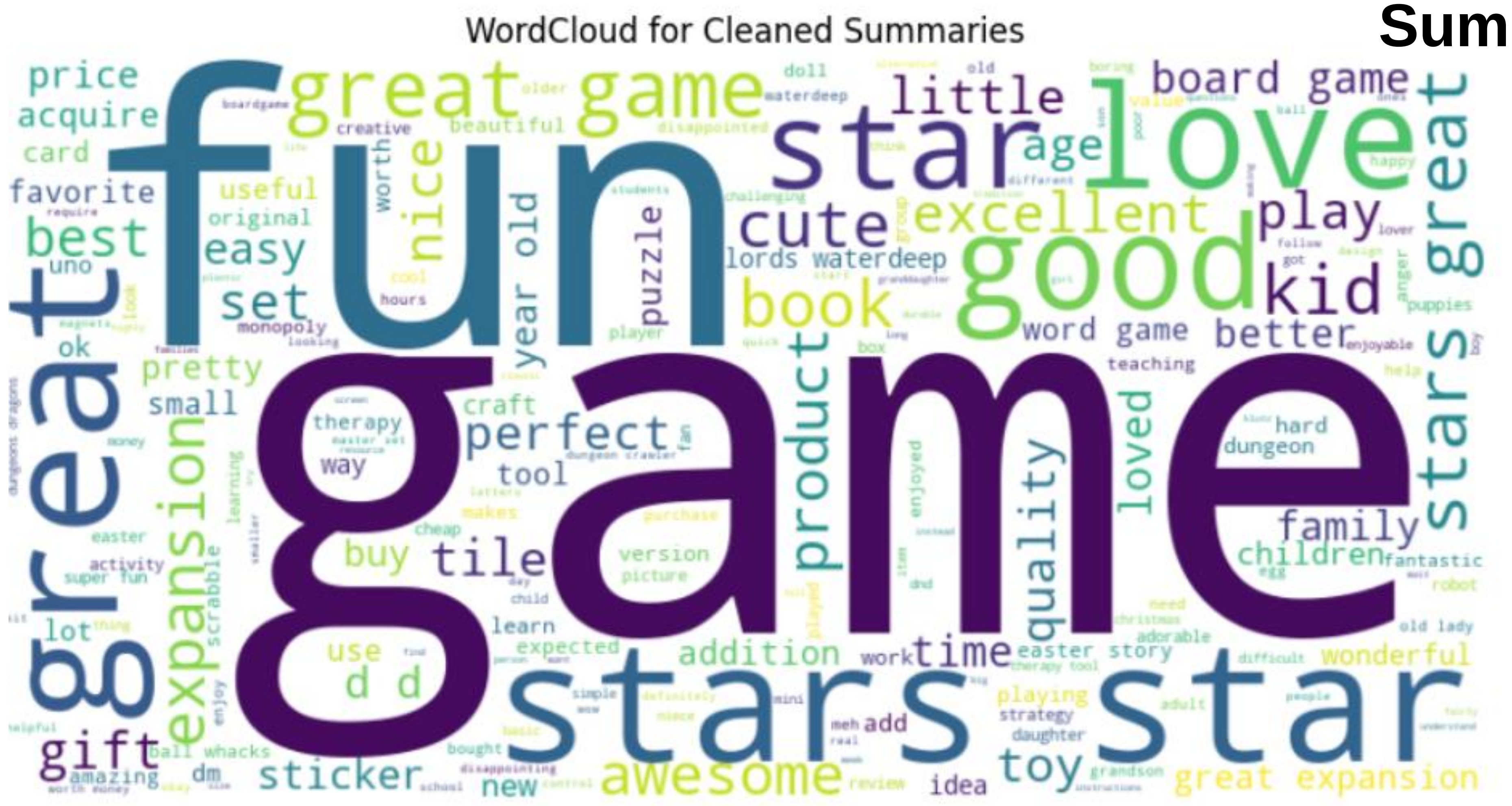
Recommendations:

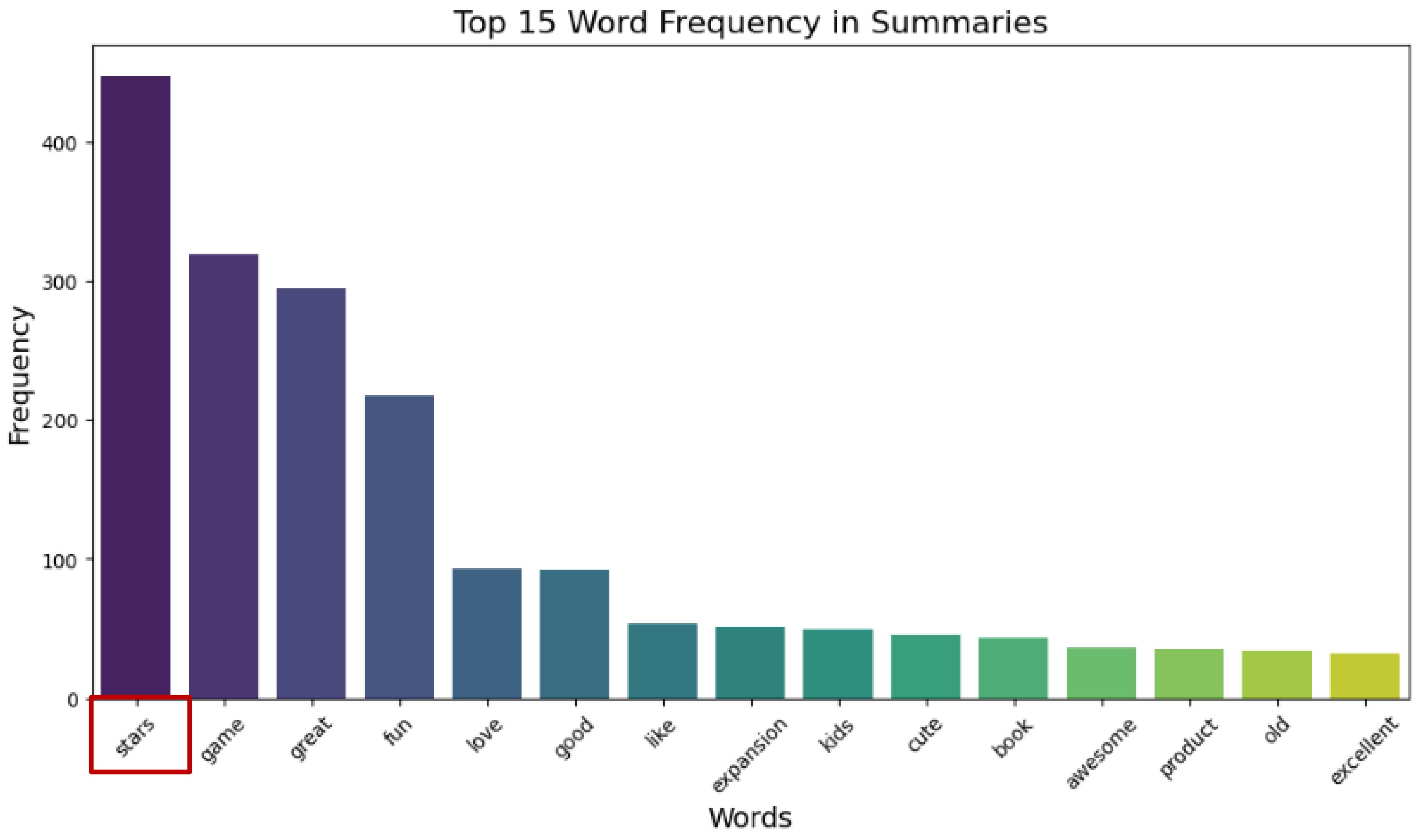
- Continue to use both methods for added benefits
- Obtain star ratings from customers besides word reviews for more accurate benchmarking

Reviews



Summaries

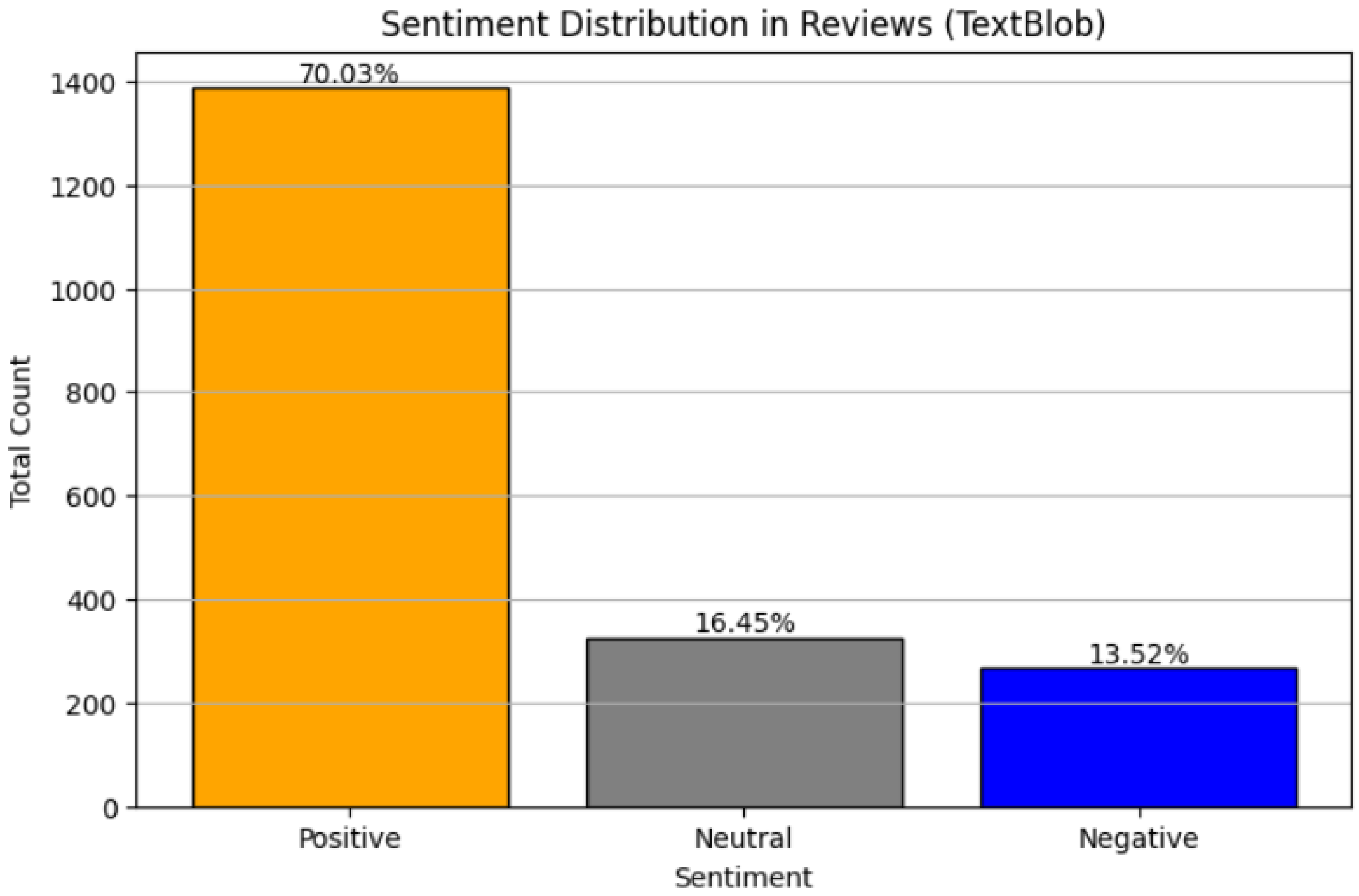


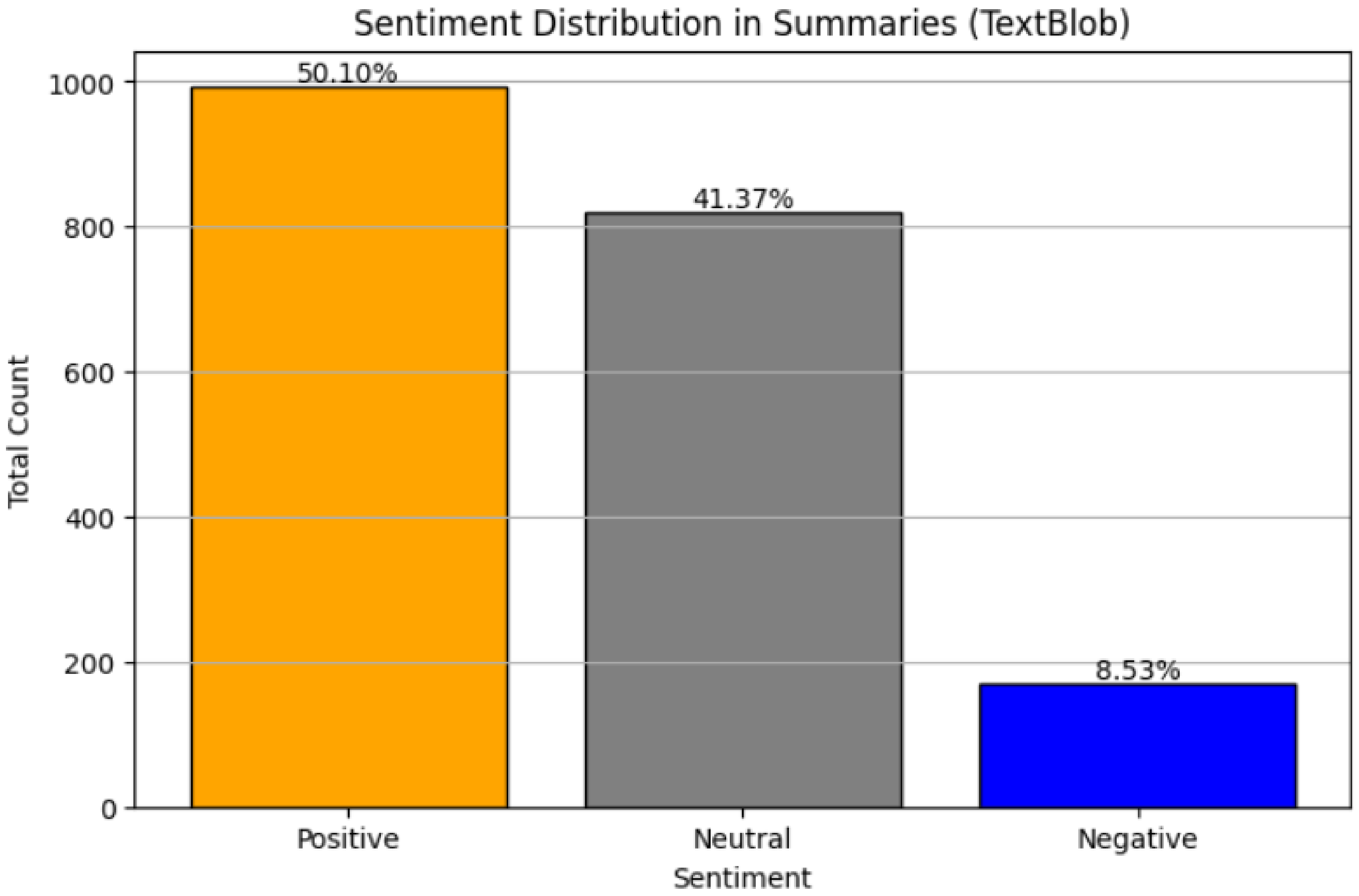


Overall Sentiment of reviews and summaries

Overall Customer Satisfaction Score (CSAT) = 78%

Note: Weighted average calculation from the different sentiment analysis CSAT results



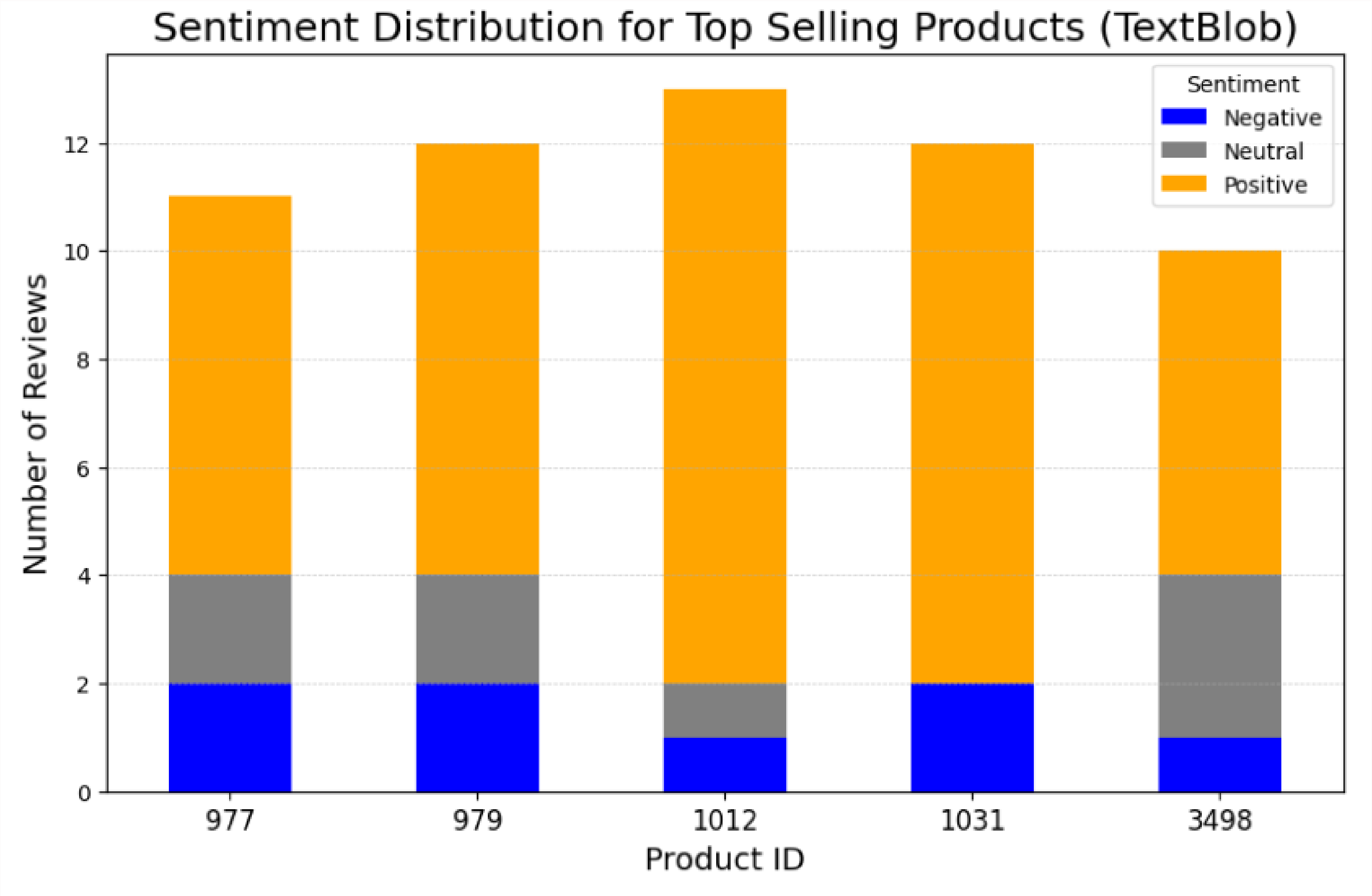


Recommendations:

- Incentivise customers to leave long reviews for continued insights for product/service improvements
- Address negative feedback promptly
- Monitor customer sentiment over time

Product related insights

- Top Products with **Positive Sentiment** (Books, Toys, Ball of Whack)
- High Engagement with **Neutral Sentiment** (Games, D&D, Monopoly)
- **Low Sentiment** Products (GaleForce9, UNO)



'Top Negative words:
'zero', 'hated', 'worthless', 'worst', 'terribly'

Continue seasonal campaigns for ***Christmas*** **and** ***Easter***

They yield strong positive engagement.

Recommendation:

- Enhance future analysis by supplementing product list and sales/profitability information
- Address Negative sentiment (e.g., with specific product feedback loops or enhanced marketing strategies)
- Refer to the report for the list of products with positive and negative reviews

Overall Business Impact

Customer Satisfaction

Customer Satisfaction is healthy but can improve, especially in product categories with neutral or negative sentiments.

Data Driven Solutions

Leveraging **predictive modelling**, **sentiment analysis**, and **demographic insights** can drive personalized marketing and customer engagement strategies.

Improve ROI metrics

Seasonal campaigns and **product-specific promotions** based on sentiment trends and customer segment preferences will continue to yield strong ROI.

Customer loyalty

Target **medium-income customers**, **incentivize spending**, and continuously **monitor sentiment** to improve overall customer loyalty.
Cheris the current “VIP”/ High loyalty customers with Personalised offerings and prime promotions.

Future Recommendations

Expand Data Collection

Consider gathering **more granular data on customer preferences, purchase history**, and interaction across platforms to refine customer segments and sentiment analysis.

Real-time sentiment tracking

Implement tools **to monitor real-time customer sentiment** for quicker feedback and timely responses.

A/B Testing for Campaigns

Run targeted A/B tests to **validate the impact of personalized campaigns** based on sentiment and customer segmentation.

Advanced Machine Learning Models

Leverage deep learning or other advanced machine learning models **to further improve predictive accuracy** for customer behaviour and sentiment.

TURTLE GAMES

GAMES THAT UNITE FAMILIES

THANK YOU

For watching this presentation